

6.2 Klasyfikacja liniowa

Zaimplementuj **analityczną** wersję klasyfikacji liniowej z regularyzacją Tichonowa(3) i wytrenuj ją na **syntetycznych jednomodowych zbiorach** danych. Porównaj na obu zbiorach działanie własnej implementacji z `sklearn.linear_model.RidgeClassifier`.

Wykorzystując zbiór danych o chorobach serca z Cleveland dokonaj eksploracyjnej analizy danych w oparciu o 13 głównych cech. Policz ilość brakujących wartości a następnie usuń wiersze je zawierające. Zastąp zmienną przewidywaną wartością binarną opisującą występowanie choroby serca. Dla wszystkich cech policz podstawowe miary statystyczne (średnia/dominanta, odchylenie standardowe, minimum, maksimum) uwzględniając typ cechy (kategoryczna/dyskretna/ciągła). Przedstaw wartości cech w formie histogramów pokolorowanych zależnie od wartości przewidywanej. Narysuj macierz korelacji pomiędzy wartościami. Które 4 cechy pozwolą na najskuteczniejszą klasyfikację?

Porównaj skuteczność własnej implementacji i `RidgeClassifier` na zbiorze danych o chorobach serca wykorzystując **13 głównych** cech. Następnie w oparciu o `RidgeClassifier` z `sklearn` i `sklearn.feature_selection.SequentialFeatureSelector` wybierz 4 najlepsze cechy. Porównaj automatyczny wybór cech do ręcznego, czy wybrane zostały inne cechy? Porównaj skuteczność własnej implementacji i `RidgeClassifier` na obu wybranych zestawach cech.

6.2.1 Kryteria oceniania

Porównanie `sklearn` i własnej implementacji klasyfikatora na syntetycznych zbiorach → 3

Eksploracja danych i uzasadniona odpowiedź na pytanie → 4

Porównanie `sklearn` i własnej implementacji na zbiorze o chorobach serca i różnych zestawach cech → 5